Crystal structure of human chorionic gonadotropin

A. J. Lapthorn, D. C. Harris, A. Littlejohn, J. W. Lustbader*, R. E. Canfield*, K. J. Machin†, F. J. Morgan‡ & N. W. Isaacs§

Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK
* Department of Medicine, Columbia University, New York 10032, USA
† St Vincent's Institute of Medical Research, Melbourne, Victoria 3065, Australia
‡ Department of Biochemistry, LaTrobe University, Melbourne, Victoria 3083, Australia

The three-dimensional structure of human chorionic gonadotropin shows that each of its two different subunits has a similar topology, with three disulphide bonds forming a cystine knot. This same folding motif is found in some protein growth factors. The heterodimer is stabilized by a segment of the β-subunit which wraps around the α-subunit and is covalently linked like a seat belt by the disulphide Cys 26–Cys 110. This extraordinary feature appears to be essential not only for the association of these heterodimers but also for receptor binding by the glycoprotein hormones.

IN humans the early stages of pregnancy are maintained by the hormone chorionic gonadotropin (hCG) from the conceptus. This hormone is structurally related to the anterior pituitary gonadotropins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which regulate the cellular and endocrine function of the ovary and testis. Together with thyroid-stimulating hormone (TSH) they constitute the family of glycoprotein hormones. These consist of two subunits, α and β, which associate non-covalently to form a heterodimer; in a given species the α-subunits are identical and the β-subunits are different (but homologous) for the different hormones. Although the heterodimer is required for receptor binding, it is the β-subunit that determines the specific activity of each hormone. The hormones are all glycosylated with *N*-linked complex carbohydrates, which confer heterogeneity on a given hormone; hCG is the most heavily glycosylated with four additional *O*-linked carbohydrates on the serine-rich C-terminal extension of the β-subunit. In human glycoprotein hormones, the common α-subunit contains 92 amino acids, with 10 half-cystine residues which form five intramolecular disulphide linkages. The β-subunits vary in size from 114 residues in LH to 145 in CG and contain 12 half-cystines which form six conserved disulphide bridges. There is a high degree of sequence similarity in the first 114 amino acids between hCG and the other hormones (LH 85%; FSH 36%; TSH 46%). The homology between hCG and LH reflects a common biological function, as both proteins bind the same receptor; FSH and TSH bind to structurally similar but distinct receptors.

Here we present the crystal structure of hCG, report the correct disulphide pairings in both subunits, and describe the overall protein fold and the dimer association that involves a structurally unique 'seat-belt' arrangement. The receptor-binding site is described. We show that both subunits of glycoprotein hormones are members of a structural superfamily of cystine-knot growth factors.

Structure determination and refinement

It was previously thought that heterogeneity of carbohydrate prevented any of the glycoprotein hormones from crystallizing. The bulk of the carbohydrate can be removed from hCG by treatment with anhydrous hydrofluoric acid, leaving the deglycosylated protein with the ability still to bind receptors, although post-translational effects are not triggered[1]. Circular dichroism shows that the structures of the deglycosylated and native hormones are identical[2]. Using the crystallization conditions for HF-treated hCG (HF-hCG)[3], isomorphous crystals of neuraminidase-treated hCG have been grown in which only the terminal sialic acid residues have been removed[4].

HF-treated hCG crystallizes from ammonium sulphate solutions[3] as hexagonal bipyramids of space group $P6_522$ and cell dimensions $a = 88.68$, $c = 177.24$ Å. The crystals diffract to 3.5 Å resolution on a laboratory X-ray source and to 2.5 Å with synchrotron radiation, but are very sensitive to radiation damage. The structure has been determined to 3.0 Å resolution by multiple isomorphous replacement (MIR) and maximum entropy (ME) solvent flattening; Table 1 gives the details and statistics for the phasing procedure (details of the ME phasing procedure will be published elsewhere).

The final electron density map allows a tracing of the α-subunit from residues 5 to 89 and the β-subunit from 2 to 111. The residue β111 is open to a large 50 Å diameter solvent channel and it is assumed that the remaining 34 C-terminal residues adopt a random conformation and are therefore not visible in the map.

The initial model was refined using the program XPLOR[5]. Five rounds of refinement and model building gave a final structure with a crystallographic *R* factor of 21.8% for data from 10–3 Å. No solvent molecules, but six *N*-acetylglucosamine molecules positioned in good density have been included. A geometric analysis of the refined structure with the program PROCHECK[6] gave values better than expected for structures determined at this resolution.

The disulphide bridges

There are five disulphide bonds in the α-subunit and six in the β-subunit. The chemical assignment of the disulphide pairings has been extensively studied (reviewed in ref. 2). The assignments of Mise and Bahl[7,8] are considered the most reliable, with α7–31, α10–32, β93–100, β26–110 and perhaps β23–72 assumed to be correct. From the ME phased map, some electron density could be interpreted as disulphide linkages 23–72, 26–110 and 34–88 of the β-subunit. The remaining S–S bridges could not be formed as expected (that is, between residues 38–57, 9–90) and the map could be interpreted only if the linkages were from residues 9 to 57 and 38 to 90, thereby forming an unusual knot

§ To whom correspondence should be addressed.

of three disulphide bridges. The density for the α-subunit contained a similar cystine knot, and for the amino-acid sequence to fit, disulphide linkages 7–31, 10–60, 28–82, 32–84 and 59–87 were necessary.

This cystine-knot motif[9] is found in three growth factors: NGF[10], TGF-β2[11] and PDGF-BB[12]. The motif involves three disulphide bridges arranged so that two disulphides link adjacent antiparallel strands of the peptide chain and form a ring penetrated by the third disulphide (Fig. 1). The disulphide pairings are: α(7–31, 10–60, 28–82, 32–84, 59–87) and β(9–57, 23–72, 26–110, 34–88, 38–90, 93–100) (Figs 1 and 2). Of the linkages not previously predicted (α10–60, 28–82, 32–84; β9–57, 38–90), all are involved in the cystine knots.

Structure of the protein subunits

Figure 2*a* shows that the subunits have similar folds determined largely by the central cystine knot. On one side of the knot there is a loop of double-stranded β-sheet-like structure (the long loop); on the other side there are two hairpin loops lying in almost parallel planes. In the β-subunit, these hairpins are linked by the Cys 23–72 disulphide bond.

In the C-terminal hairpin loops of each subunit, there are a pair of β-sheet bulges at adjacent positions in each strand (α62–65, 79–82; β59–62, 85–88). In the β-subunit, the bulges have a classical hydrogen-bonding pattern; in the α-subunit the pattern is less complete. The structures of all the cystine-knot growth factors have β-sheet bulges located immediately after Cys(V), suggesting they are an integral component of this motif.

The long loop in the α-subunit consists of a stretch of antiparallel β-sheet made up of residues 35–39 and 52–56. The three residues Pro38-Thr39-Pro40, which are conserved in all 18 α-subunit sequences determined, break the β-sheet away at an angle of about 100° and lead into two turns of a 3_{10}-helix involving residues 40–46. Asn 52 is also on a bend that directs the *N*-linked carbohydrate on this residue out from the interface between the α- and β-chains.

The corresponding region of the β-subunit forms a very open loop with no main-chain hydrogen bonds between the antiparal-

TABLE 1 Summary of crystallographic analysis

Diffraction data							
Data set	Native	$K_2Pt(CN)_4$	$AgNO_3$	$AgNO_3$	$Hg(OOCH_3)_2$	$KAu(CN)_4$	$K_2Pt(CN)_4$
Detector	Xen./Film	Xen.	Xen.	Xen.	Film	Film	Film
Resolution (Å)	3.0	3.5	5.0	3.5	3.5	3.9	3.3
Unique reflections	8,365	3,169	1,096	3,664	2,473	2,272	2,949
Completeness (%)	94.9	58.9	81.6	70.6	44.0	55.2	52.4
R_{merge} (%)	11.1	6.0	7.7	7.5	10.0	9.1	7.5
MIR phasing							
Mean fractional ID		0.126	0.094	0.115	0.080	0.148	0.073
Number of sites		2	2	3	1	2	1
Rc		0.73	0.68	0.82	0.80	0.80	0.78
Phasing power centric		1.1	1.0	0.7	0.7	0.7	0.6
acentric		1.4	1.5	1.1	0.9	1.0	0.6
Phase combined FOM	0.52						

Crystallization and data collection: Crystals were grown at 25 °C as described[3]. More than 20 heavy-atom compounds were tested with a variety of soak conditions to yield four useful derivatives. Diffraction data were collected on film on station 7.2 (λ=1.48 Å) at the SERC Synchrotron Radiation Source, Daresbury. Film data were processed using the MOSFLM program and native and derivative data were scaled using SCALEIT from the CCP4 suite of programs[40]. Area detector data were collected on a Siemens/Xentronics P1000 area detector, and processed using the XDS program[41]. The final native dataset used for refinement, on which the current model is based, is a combination of a complete Xentronics dataset to 3.5 Å resolution collected from a single crystal, merged with the less complete film data collected from several crystals which extends to 2.8 Å. From a consideration of the merging *R* factor statistics, only data to 3.0 Å were used. XEN, Xenotronics area detector; ID, isomorphous differences. $R_{merge} = \Sigma|I_i - \langle I\rangle|/\Sigma I_i$, where I_i is the intensity of an individual reflection and $\langle I\rangle$ is the mean intensity of that reflection; **MIR analysis:** The Pt derivative was solved using SHELXS86[42] and the other derivatives by a combination of difference Fourier syntheses and difference Patterson maps. There are essentially three sites; the major site shared by the Pt and Au derivatives is in a pocket on a 2-fold axis of symmetry binding β45 Arg and its 2-fold symmetry equivalent. Both the primary Ag site and the only Hg site are very close (sandwiched between α29 Met and β41 Met) and the other Ag sites are at α83 His and α71 Met. The heavy-atom parameters were refined and MIR phases calculated using MLPHARE for the six derivative datasets, using both isomorphous and anomolous differences to give a FOM of 0.52. The correct space group was determined from refinement of heavy-atom sites in MLPHARE using anomalous data from the Pt derivative and confirmed by the handedness in the twist of β-sheets in the final structure. Cullis *R*-factor, $Rc = \Sigma||F_{PH} \pm F_P| - F_H|/\Sigma|F_{PH} - F_P|$ where F_P and F_{PH} are the structure factor amplitudes for native and derivative data respectively and F_H is the calculated heavy-atom structure factor; phasing power $= \langle F_H\rangle/\langle E\rangle$, where $\langle F_H\rangle$ is the r.m.s. of the heavy-atom scattering factor and $\langle E\rangle$ is the r.m.s. lack of closure; the summation is computed for reflections used in the heavy-atom refinement cycle. FOM, figure of merit. **Density modification and model building:** The initial phases were refined at 3.5 Å resolution using the Wang solvent-flattening procedure[43]. A protocol of five cycles of solvent flattening following an envelope redetermination (55–60% solvent) was used for 10 rounds of phase refinement. Phase extension from 4.0 Å using very small increments in resolution was used in an attempt to improve the quality of the maps. Although ~75% of the protein chain, including the major β-strands in the structure, could be traced, the maps were not sufficiently clear to define the subunit interface or to get a sequence alignment. Further density modification was achieved using the program MICE[44]. Using MIR-phased reflections with FOM $\geq$0.70 as basis-set reflections and all data to 3.0 Å resolution, maximum entropy solvent flattening (MESF) was performed. Seven rounds of MESF were calculated with phase recombination computed using SIGMAA[45] and envelope determination using the Wang procedure. A round of phase permutation using 7 important unphased reflections gave phases for 5 of these, which were then combined with the initial MIR basis-set reflections and an additional seven rounds of MESF computed. Partial structure phasing from 14 peptide fragments, with side chains assigned according to size was included in phase recombination in the later rounds of MESF. This led to a map which was sufficiently unambiguous to allow the complete tracing of both subunits. **Refinement:** The polypeptide chain was built into a combined map using the programs FRODO[46] and O[47]. This model was refined at 3.5 Å using the program XPLOR[5]. Five rounds of model rebuilding and refinement using both conventional positional refinement and simulated annealing, during which the resolution was extended to 3.0 Å, resulted in the current model. The model comprises a total of 1,565 non-hydrogen atoms. Temperature factors were refined individually and then averaged by group with a final overall *B* value of 26.5 Å^2. Data from 10 to 3.0 Å with $F > 2\sigma(F)$ (7,877 reflections) were used in the final refinement to give an *R* factor of 21.8%, and a Free *R* factor (791 reflections) of 31.4%. The stereochemistry is typified by an r.m.s. bond deviation from ideality of 0.017 Å and r.m.s. angles of 2.28°. All but one of the non-glycine ψ and ϕ angles lie in allowed regions in the Ramachandran plot, with 77% in the most favoured regions. Atomic coordinates will be deposited in the Brookhaven Protein Data Bank.

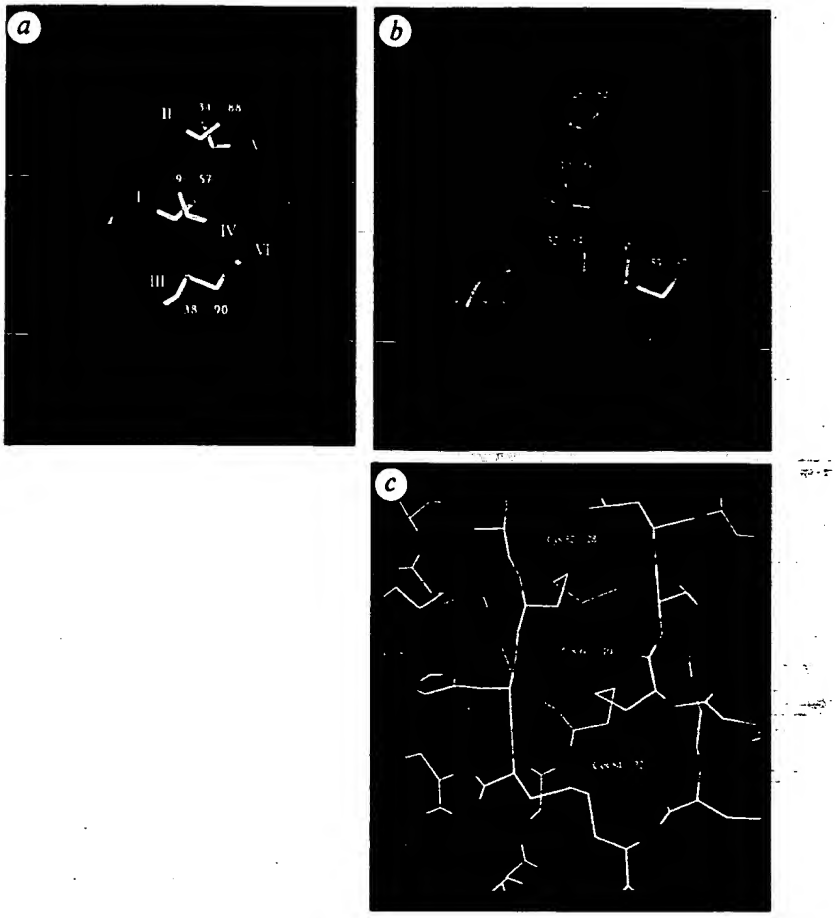


FIG. 1 The structure of the α- and β-subunit cystine knots shows the common fold that comprises three disulphide bridges, arranged so that two disulphides link adjacent parallel strands of the peptide chain and form a ring through which the third disulphide penetrates. *a*, The common arrangement of the half cystines in the sequence[9] is shown for the β-subunit. If the half cystines within the knot are assigned roman numerals to indicate their order in the protein sequence, disulphide bonds Cys(II)–Cys(V) and Cys(III)–Cys(VI) form the ring through which the interpenetrating disulphide Cys(I)–Cys(IV) is formed. There are two general sequence requirements, Cys(II)-X-Gly-X-Cys(III) and Cys(V)-X-Cys(VI), although this is not absolute as in NGF, where the central Gly in the first motif is replaced by a 7-amino-acid loop (Fig. 5). These sequence patterns are found in each subunit of hCG: α28–32 (CMGCC), α82–84 (CHC), β34–38 (CAGYC) and β88–90 (CQC). In the isolated subunits, the disulphides in the cystine knots show varying degrees of solvent accessibility, as predicted[48]. Cys(I)–(IV) and Cys(II)–Cys(V) are buried in the knot, but Cys(III)–Cys(VI) is exposed to the solvent like the remaining disulphide bridges. *b*, In the α-subunit there are additional disulphide bridges (7–31 and 59–87) in close proximity to the knot. These invariant residues seem to play a role in formation of the heterodimer. There is a notable similarity in overall structure in the knots and the only significant differences, such as the dihedral angle of Cys(III)–Cys(VI), are due to conformational restraints imposed to accommodate the additional cystines in α. *c*, The $(2|F_{obs}| - |F_{calc}|)$ electron density map around the cystine knot of the α-subunit is indicative of the quality of the final map. The refined structure of the protein is illustrated as an atomic stick figure, colour coded for atom type: C, white; O, red; N, blue; S, yellow. The electron density is contoured at a 1σ level.

lel strands. Instead, the loop is stabilized by side chain-to-main chain hydrogen-bond interactions from Arg 43 to Pro 50 and Leu 52 and from Gln 54 to Met 41. One side of the loop, residues 38 to 45, is held by association with the α-subunit. In the crystal, a two-fold axis of symmetry causes the hydrophobic residues 48–50 to pack against the same residues from a neighbouring molecule.

Apart from the cystine-knot motif, there is no sequence similarity between the subunits, but when the cystine knots are superimposed, the similarity in structure is remarkable (Fig. 2*b*). The

FIG. 2 The structure of each subunit is shown as a Cα trace with disulphide bridges. (The α-subunit is blue and the β-subunit is green in this and all subsequent figures.) Each molecule has a similar fold consisting of two β-hairpin loops on one side of a central cystine knot and a long loop on the other. *a*, The hairpin structure of each subunit is stabilized by a hydrophobic core extending between the two loops. The α-subunit core consists of α17 Phe, α18 Phe, α25 Ile, α68 Val, α70 Val, α71 Met, α74 Phe and α76 Val, with the three phenylalanines clustered together and partially exposed to the surface. The β-subunit core includes the residues β16 Leu, β18 Val, β27 Ile, β29 Val, β67 Ile, β69 Leu, β80 Val and β82 Tyr, which partly buries the disulphide bridge β23–72. The two hairpins of α are essentially regular, apart from an unusual bend at the invariant Asn 15 where the side chain hydrogen-bonds to the main chain NH of residues 17 and 18, resulting in Pro 16, Phe 17 and Phe 18 bulging away from the antiparallel strand and forming a prominent surface. Residues 90–111 are highlighted in red on the β-subunit. This 'seat-belt' region is a novel structure essential for forming the heterodimer. *b*, Superimposing the α- and β-subunits at the cystine knot reveals the extent of their common structure. The Cα position of 22 residues extending from the knot can be superimposed with a r.m.s. difference of 0.52 Å. Subunits are coloured as in *a*, with α-subunit disulphides in orange.

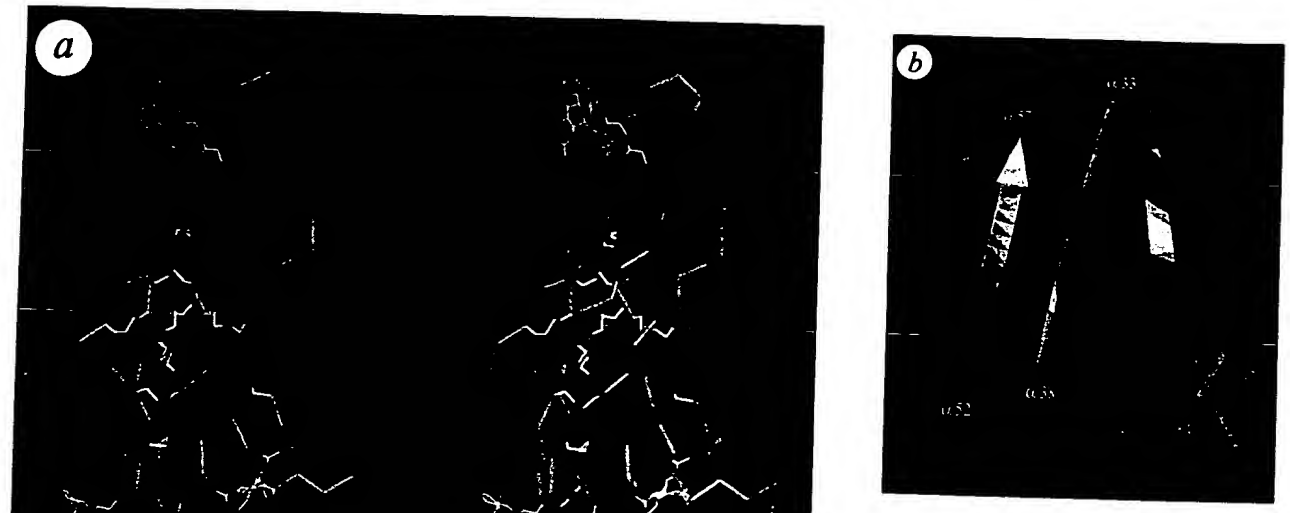


FIG. 3 *a*, A stereo view of the complex of α- and β-subunits (blue and green, respectively) showing the disulphide bridges (yellow) and observed carbohydrate (using colour coding on atom type). *b*, A schematic representation of the central β-barrel. One face consists of a short strand from the β-subunit (97–102) forming a parallel β-sheet with α(52–57), which is part of a four stranded antiparallel β-sheet as shown. The other face of the barrel has two antiparallel strands (84–90, 63–56) of the β-subunit.

two molecules have a common core structure and differ mainly in the extent of the loop regions. In the β-subunit the pair of hairpin loops are extended in size, while the long loop is larger in the α-subunit.

Formation of the heterodimer

Figure 3*a* shows the heterodimer. The two subunits are related by an approximate two-fold axis. Segments of well defined β-sheet structure near the cystine knot in each subunit are brought together to form a short seven-stranded β-barrel (Fig. 3*b*). The β34–36 (CAGY sequence), long implicated in the formation of the dimer[2], is in the central strand. Away from this central region, there are a number of hydrogen bonds (data not shown). There is a short stretch of antiparallel β-sheet (β44–46 and α77–75) which helps to hold the long loop of the β- subunit to the hairpin loops of the α-subunit. In addition, there are hydrophobic contacts between valine and leucine at β44–45 and the triplet of phenylalanine residues (α17, 18, 74) in the hairpin loop of the α-subunit. The complete association buries a total surface area of 4,525 Å^2 between the subunits: 2,134 Å^2 by α 2,241 Å^2 by β, and 150 Å^2 by the carbohydrate on the α-subunit, specifically that of Asn 52.

An unexpected feature of the association is the role played by the loop from Cys 90 to Cys 110 in the β-subunit (Fig. 2*a*). In the heterodimer this loop is wrapped over the α-subunit while remaining covalently bonded to the β-subunit through the disulphide linkages (9–90, 26–110), and has the appearance of a seat belt (Fig. 3*a*). There is close association between the inner surface of the belt and the α-subunit with a short strand of parallel β-sheet between β(99–101) and α(53–57) forming part of the central β-barrel. The side chains of four α-subunit residues form a surface which interacts with either main chain or side chain atoms of the belt. There are hydrogen bonds between the side chains of α35 Arg and β106 His and the side chain of α56 Glu and the main chain of β104 Lys. Non-bonded interactions are formed between α37 Tyr, β107 Pro and β108 Leu and between α54 Thr and β99 Asp. There are no contacts specific to the hCG molecule that would prevent this seat-belt arrangement existing in all of the glycoprotein hormones.

Formation of Cys 93–100 is necessary for association with the α-subunit[13] and a form of hCGβ consisting of residues 8–100 associates only weakly with α[14]. The disulphide Cys 26–110 is completed after the α/β association has occurred[15], indicating that the seat-belt region is important in maintaining the integrity of the heterodimer.

Carbohydrate structure

In its native form, hCG is heavily glycosylated with complex carbohydrate making up ~34% by weight of the protein. *N*-linked carbohydrates occur at Asn 52 and Asn 78 on the α-subunit and at Asn 13 and Asn 30 on the β-subunit. In addition, the β-subunit has four *O*-linked carbohydrates at Ser 121, 127, 132 and 138. Treatment of the protein with anhydrous HF for one hour leaves the *O*-linked carbohydrate largely intact and truncates the asparagine-linked carbohydrate to predominantly Asn-$(GlcNAc)_2$ (refs 4, 16).

In the crystal structure, electron density is clear for two sugar residues on each of the α-subunit *N*-linked carbohydrates (Fig. 4*a*) and only one sugar residue is seen for those of the β-subunit. These fragments are sufficient to allow modelling of the complete carbohydrate (Fig. 4*b*). In the α-subunit, the carbohydrates are at the extremities of the molecule: Asn 52 is on the double-stranded loop and Asn 78 is near the end of the β-hairpins. The carbohydrate structure makes few hydrogen bonds to the protein. In the β-subunit, both *N*-linked carbohydrates are exposed on the outward faces of adjacent β-strands with Asn 13 and Asn 30 less than 7 Å apart (Fig. 3*a*). Some of the glycoprotein hormones have only one of these sites glycosylated. The only carbohydrate that interacts at the subunit interface is on α52 Asn which contacts the β-subunit residues

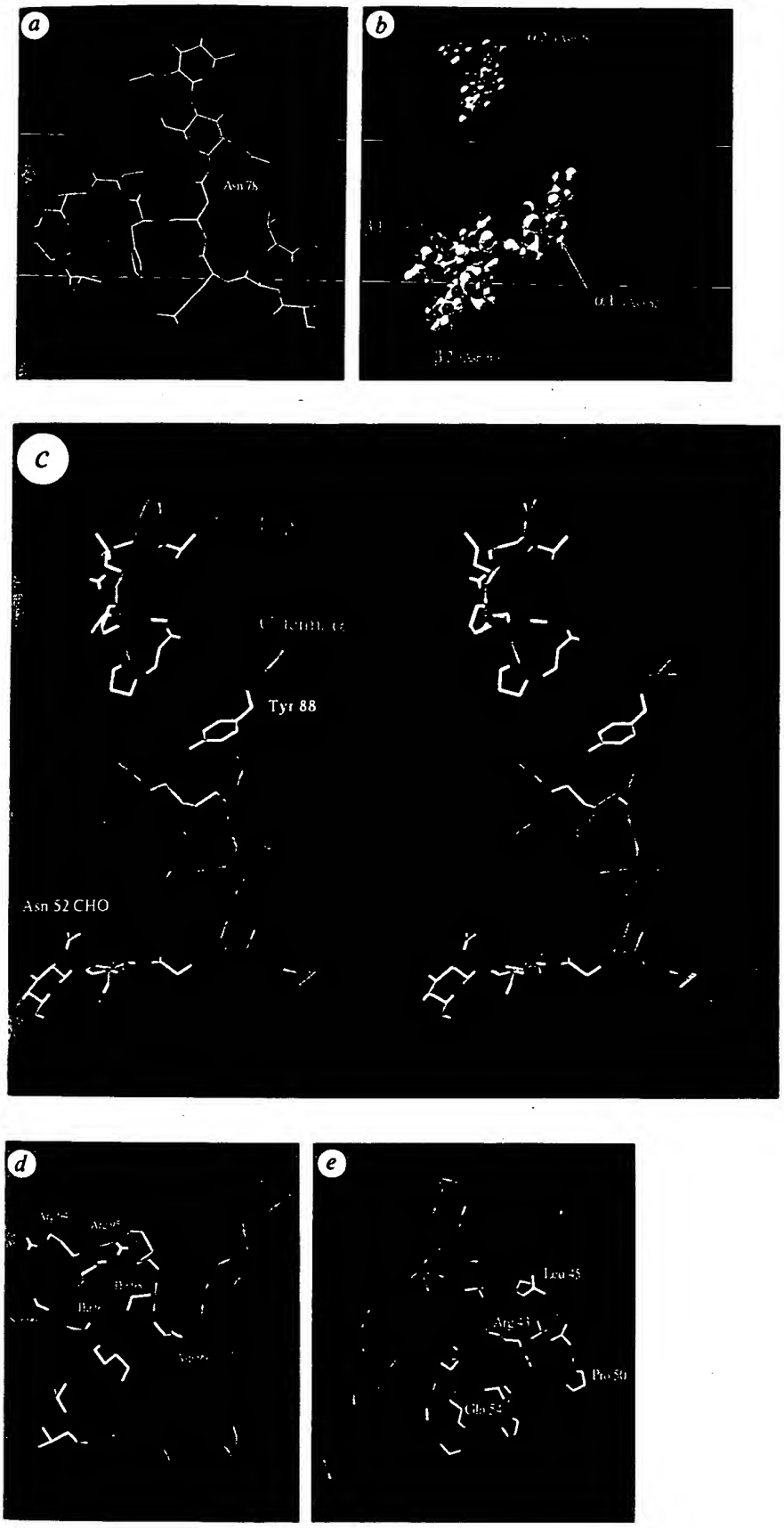


FIG. 4 *a*, Carbohydrate density at α78 Asn contoured at a 1.0σ level in the final $2|Fo-Fc|$ difference Fourier, showing clear density for two *N*-acetyl-glucosamine units. *b*, A model of the complete *N*-linked carbohydrate on the hCG molecule illustrates the extent of the carbohydrate structure. The protein is represented as surfaces[49] (α, blue; β, green), with the carbohydrate represented as a space-filling model with atom colouring: N, blue; O, red; C, white. The carbohydrate was modelled by optimizing the fit of a complex biantennary carbohydrate from thrombin (coordinates provided by P. Martin) on the observed sugar residues at each of the glycosylation sites. The carbohydrate at α78 is far removed from the other three sites. The positions of the β13 and β30 carbohydrates are very close, forming a large bulk of carbohydrate. Some distal residues of the β carbohydrate could be close to one of the antennae of the α52 carbohydrate which is located at the dimer interface and is within the proposed receptor-binding domain. This sugar is best placed for biological function. *c*, A stereo view of the proposed receptor-binding site. The essential determinant loop (β93–100, residues in red) is centrally positioned with the β-subunit long loop (38–57, coloured by atom type) and Tyr 88 (white) of the α-subunit C-terminus above. Below is the carbohydrate (coloured by atom type) at α52 Asn, with the adjacent α40–50 loop. Some invariant residues are coloured magenta. *d*, Side chains in the determinant loop β93–100, with colour coding for atom type. The adjacent helical region of the α-subunit shows the position of Leu 48, Val 49 and Lys 51. *e*, The unusual wedge-shaped β38–57 long loop extends from the hairpin loops of the α-subunit. Hydrophobic residues are predominantly surface-oriented; hydrophilic residues are either in the centre of the loop or buried in the α/β interface.

Tyr 59, Val 62, Phe 64 and Ala 83 as well as Thr 97, from the determinant loop.

Receptor binding and biological activity

A number of residues important for the activity of hCG have been identified through chemical modification, the use of synthetic peptides in competitive inhibition studies, and site-directed mutagenesis (reviewed in refs 2, 17, 18). As a result, several regions that contribute to receptor binding have been identified (Fig. 4).

The (β93–100) determinant loop sequence (Fig. 4*d*). The importance of this sequence was recognized by Ward and Moore[19], who proposed that the variable charge in this loop could act as a determinant of specificity, positive for LH/hCG and negative for FSH and TSH. The disulphide 93–100 which forms this loop is important for activity[20] and site-directed mutagenesis of either Cys 93 or Cys 100 in hCGβ[21] yields mutant forms incapable of associating with the α-subunit.

The determinant loop is surface-oriented and held in place by a short stretch of parallel β-sheet between α53–57 and β99–101. The disulphide β93–100 sits between the α53–56 and β50–57 strands, making non-bonded contacts with α55 Ser and β56 Val and constrains the loop to form a single turn of 3_{10} helix-like structure between β94 Arg and β98 Thr. This latter residue is buried at the interface with the α-subunit and its importance in subunit association has been demonstrated by mutagenesis[22]. Residues of the β-subunit implicated in receptor binding[22] (Arg 94, Arg 95, Ser 96 and Asp 99) are accessible at the dimer surface. Thr 97, generally insensitive to mutagenesis, is in close contact with the carbohydrate of α52 Asn.

The (β38–57) sequence (Fig. 4*e*). Much attention has been focused on the longest inter-cysteine loop β38–57. Synthetic peptides with this sequence stimulate steroidogenesis in rat Leydig cells[23] and strongly inhibit binding of whole hCG[23].

The crystal structure shows an unusual loop with residues 40–46 buried at the α/β interface (Fig. 3). Mutagenesis of Arg 43 to Leu (as in FSH) in hCG[24], or replacement of it by Ala or Asp in synthetic (38–57) peptides[25] either significantly diminishes binding or eliminates it. A natural mutation in LH of Gln 54 to Arg causes hypogonadism and the mutant hormone does not bind to its receptor[26]. Residues 47–53 are predominantly hydrophobic and exposed at the surface, forming a wedge-shaped extrusion (Fig. 4*e*). This unusual structure, by virtue of its composition and proximity to the determinant loop, is likely to be important in receptor binding. The orientation of Arg 43 suggests that it may stabilize the conformation of the loop, although direct interaction with the receptor cannot be excluded.

The (α88–92) C-terminus. Three of the five C-terminal residues (Tyr-Tyr-His-Lys-Ser-COOH), are indentical in 13 of the 18 available amino-acid sequences. Carboxypeptidase digestion of both hCGα and LHα, although not detrimental to subunit assembly, essentially eliminates receptor binding[27,28]. hCGα lacking residues 89–92 forms an active heterodimer but with reduction of both receptor binding and steroidogenesis[29].

Although residues 89–92 could not be positioned, the C-terminus of the α-subunit should be located close to both the β38–57 sequence loop and the determinant loop of the β-subunit, forming a composite receptor-binding site (Fig. 4*c*). The side chain of α88 Tyr packs between the disulphides β93–100 and α59–87. Mutagenesis of this residue to Phe causes reduced binding affinity but increases the steroidogenic response at saturating concentrations[29]. As it is unlikely to be involved directly in receptor binding, this result may be due to subtle differences in conformation induced in the final four C-terminal residues.

Other potential α-subunit residues. The α-subunit residues 40–50 contain the only helical structure in the protein. The helix consists of highly conserved residues and is adjacent to the determinant loop (Fig. 4*c*, *d*), suggesting a receptor-binding role. This

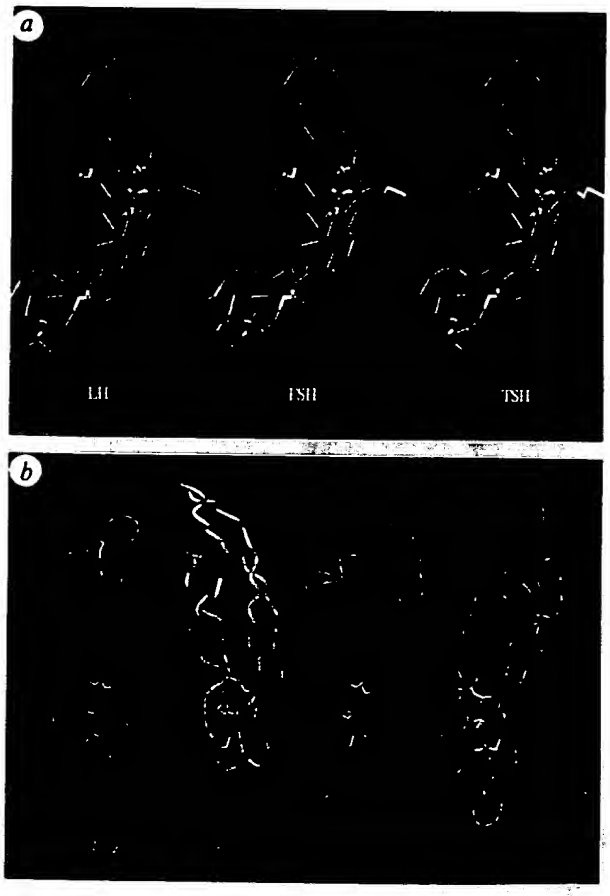


FIG. 5 *a*, A Cα trace of hCG-β, coloured according to variation in the amino-acid sequence of the other human glycoprotein hormones. A log-odds matrix[50] was used to define non-conservative (red) and conservative (cyan) sequence changes. Invariant residues are coloured green, deletions white. The conserved cystine side chains are also shown. *b*, Schematic drawings of structures of hCG-β (green) compared to PDGF-β (tan), TGF-β (magenta) and NGF (blue) (coordinates from the Brookhaven Protein Data Bank[51]; not given for residues 27–37 in PDGF-β). The N termini and C termini of hCG-β and TGF-β have been truncated to simplify the diagram. The structures share the same cystine-knot motif, imposing strong conformational restraints on associated residues. There are significant changes in the size and shape of the hairpin loops, but the largest differences are seen in the long loop. Positions where hCG-β is cleaved to form the β-core fragment are shown in red and highlight the similarity between this fragment and PDGF-β.

is supported by recent work[30] implicating Lys 44 in receptor binding. The structure further suggests that the invariant Leu 48 and Val 49, which protrude into solvent at a turn, and Lys 51, which extends up towards the determinant loop, also play a part.

Carbohydrate. As well as modifying the rate of clearance of the glycoprotein hormone from the body, some glycosylation is essential for full biological activity[31]. The LH/CG receptor contains, within the extracellular domain, a sequence similarity to soybean lectin[32], which suggests that there is a binding site for carbohydrate. HF-hCG binds to receptors with slightly increased affinity but with loss or reduction of potency[33]. Studies on the function of the carbohydrate show that the *N*-glycosylation sites are more important than the *O*-linked sites; also, glycosylation is not necessary for receptor recognition, and *N*-glycosylation of the α-subunit at Asn 52 is critical for signal transduction[34,35].

The carbohydrate at α52 Asn is positioned at the dimer interface, about 4 Å away from the determinant loop (Fig. 4*c*). Its position, size and flexibility could be antagonistic to a peptide-based receptor binding, thereby explaining how removing carbohydrate increases receptor binding. Removal of both β-subunit *N*-glycosylation sites does not cause much loss of activity, but reduces the maximal steroidogenic activity[36]. Although the two Asn residues are very close to each other and located 20 Å from α52 Asn, modelling with the complete complex carbohydrate shows that some of the distal sugar residues on β13, β30 and α52 could be close together (Fig. 4*b*).

Deglycosylated hCG can be converted back into an active form when bound to polyclonal antibodies[31,37], suggesting that a modified conformation is restored to the native form by the bound antibody[31]. As crystals of asialo-hCG are isomorphous with HF-hCG, there is no evidence in the crystal structure for conformational change caused by deglycosylation. The structure does support the alternative explanation[37] that the antibodies may mimic the steric interaction of the carbohydrate with the receptor.

The family of glycoprotein hormones

It is evident that the structures of the β-subunits of the other members of this family (LH, FSH, TSH) will be essentially similar to the structure described here. Figure 5*a* shows the positions of the non-conserved, conserved and invariant residues for each of the human hormones when compared to hCG. LH has only a few random non-conserved changes, as expected. FSH and TSH both have functionally distinct receptors and less sequence similarity. The non-conservative sequence changes are concentrated in the receptor-binding regions, namely the 38–57 long loop and the 93–100 determinant loop. There are unexpected changes in regions associated with the α-subunit, predominantly in the 100–110 seat belt and 39–47 region of the long loop. Conserved and invariant residues are concentrated in the two hairpin loops and the cystine knot, implying that these are important in maintaining the fold. Charged surface residues are also conserved in the hairpin loops, indicating that they may serve some common function for these hormones.

Cystine-knot growth factors

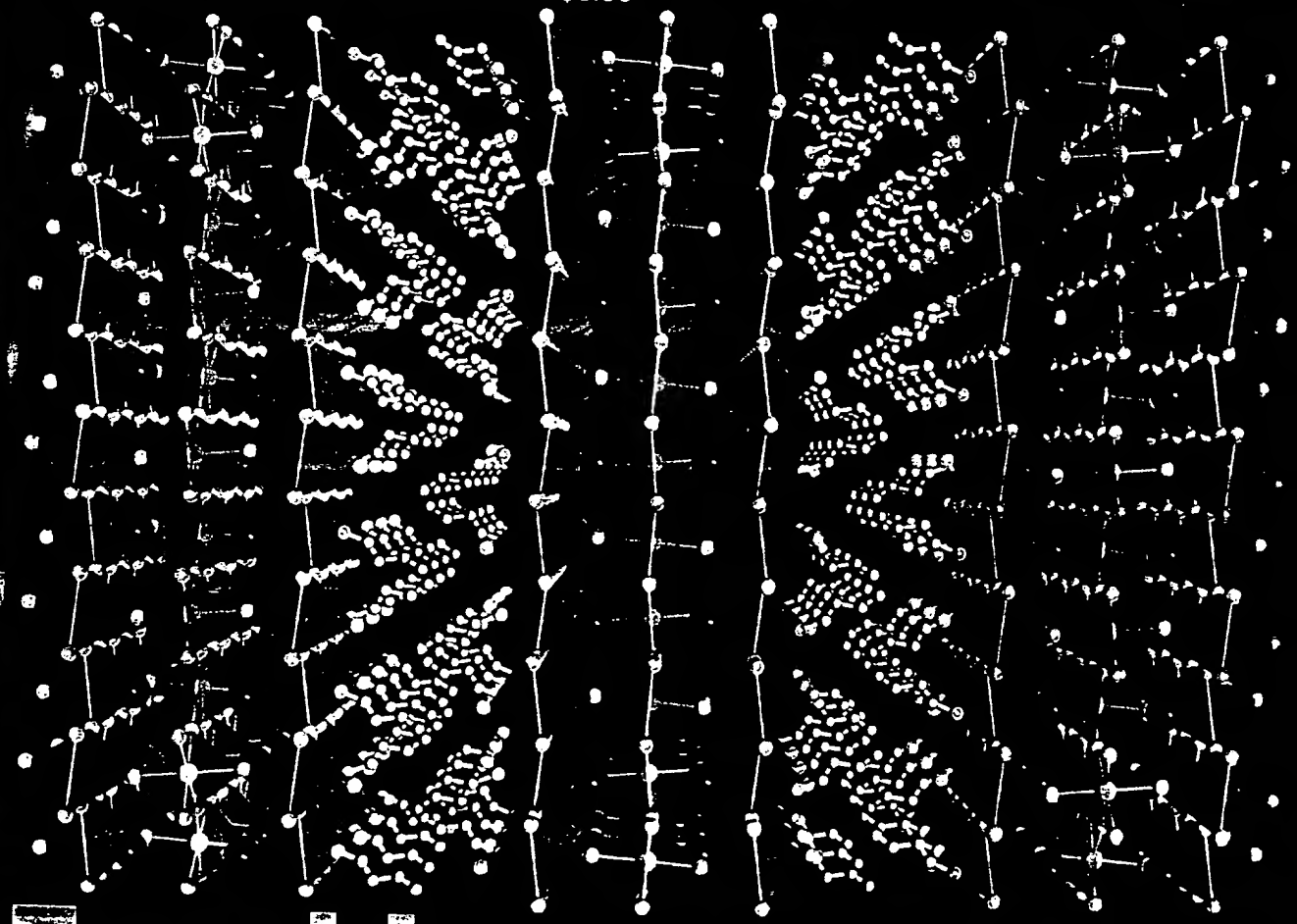
The correct assignment of the disulphide bridges and the determination of the tertiary fold of hCG reveals, surprisingly, that the glycoprotein hormones are members, with NGF, TGF-β and PDGF-β, of the superfamily of cystine-knot growth factors. In this family the proteins form either homodimers or heterodimers with homologous cystine-knot subunits. In contrast, the glycoprotein hormones form only heterodimers between a common α-subunit and different β-subunits. Structural comparisons with TGF-β, PDGF-β and NGF show that the cystine-knot motif allows variation in the size and conformation of the constituent loops, but there is a pronounced similarity between part of hCG-β and PDGF-β (Fig. 5*b*). Whereas hCG is found mainly as a heterodimer in the early stages of pregnancy, in the later stages both subunits are found in large quantities as separate molecules. Free α-subunit is more heavily glycosylated and has hormonal activity[38]. Smaller forms of the β-subunit that have been considered degradation products are found. The smallest of these, β-core[39], is composed of two polypeptides, β6–40 and β55–92, which arise from excision of 15 residues from the long loop as well as truncation at the N and C termini. The structural similarity of this truncated β-molecule with PDGF is striking and suggests that the circulating β-core molecule has a defined tertiary structure and perhaps a biological function. □

Received 2 March; accepted 12 May 1994.

1. Manjunath, P. & Sairam, M. R. *J. biol. Chem.* **257,** 7109–7115 (1982).
2. Ryan, R. J. *et al. Rec. Prog. Horm. Res.* **43,** 383–429 (1987).
3. Harris, D. C., Machin, K. J., Evin, G. M., Morgan, F. J. & Isaacs, N. W. *J. biol. Chem.* **264,** 6705–6706 (1989).
4. Lustbader, J. W. *et al. Biochemistry* **28,** 9239–9243 (1989).
5. Brunger, A. T. *XPLOR Manual* (Yale Univ. Press, New Haven, 1992).
6. Morris, A. L., Macarthur, M. W., Hutchinson, E. G. & Thornton, J. M. *Proteins* **12,** 345–364 (1992).
7. Mise, T. & Bahl, O. P. *J. biol. Chem.* **255,** 8516–8522 (1980).
8. Mise, T. & Bahl, O. P. *J. biol. Chem.* **256,** 6587–6592 (1981).
9. Murray-Rust, J. *et al. Structure* **1,** 153–159 (1993).
10. McDonald, N. Q. *et al. Nature* **354,** 411–414 (1991).
11. Schlunegger, M. P. & Grutter, M. G. *J. molec. Biol.* **231,** 445–458 (1993).
12. Oefner, C., D'Arcy, A., Winkler, F. K., Eggiman, B. & Hosang, M. *EMBO J.* **11,** 3921–3926 (1992).
13. Chen, F. & Puett, D. *J. biol. Chem.* **266,** 6904–6908 (1991).
14. Puett, D., Huang, J. & Xia, H. in *Glycoprotein Hormones* (eds Lustbader, J. W., Puett, D. & Ruddon, R. W.) 122–134 (Springer, New York, 1994).
15. Huth, J. R., Mountjoy, K., Perini, F. & Ruddon, R. W. *J. biol. Chem.* **267,** 8870–8879 (1992).
16. Keutmann, H. T., McIlroy, P. J., Bergert, E. R. & Ryan, R. J. *Biochemistry* **22,** 3067–3072 (1983).
17. Sairam, M. R. in *Hormonal Proteins and Peptides* (ed. Li, C.H.) **XI,** 1–79 (Academic, New York, 1983).
18. Keutmann, H. T. *Molec. cell. Endocr.* **86,** c1-c6 (1992).
19. Ward, D. N. & Moore, W. T. in *Animal Models for Research on Contraception and Fertility* (ed. Alexander, N. J.) 151–161 (Harper and Row, New York, 1979).
20. Keutmann, H. T., Mason, K. A., Kitzmann, K. & Ryan, R. J. *Molec. Endocr.* **3,** 526–531 (1989).
21. Suganuma, N., Matzuk, M. M. & Boime, I. *J. biol. Chem.* **264,** 19302–19307 (1989).
22. Huang, J., Ujihara, M., Xia, H., Chen, F., Yoshida, H. & Puett, D. *Molec. cell. Endocr.* **90,** 211–218 (1993).
23. Keutmann, H. T. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 2038–2042 (1987).
24. Chen, F. & Puett, D. *Biochemistry* **30,** 10171–10175 (1991).
25. Keutmann, H. T. *et al. Biochemistry* **27,** 8939–8944 (1988).
26. Weiss, J. *et al. New Engl. J. Med.* **326,** 179–183 (1992).
27. Merz, W. E. *Eur. J. Biochem.* **101,** 541–553 (1979).
28. Parsons, T. F. & Pierce, J. G. *J. biol. Chem.* **254,** 6010–6015 (1979).
29. Chen, F., Wang, Y. & Puett, D. *Molec. Endocr.* **6,** 914–919 (1992).
30. Xia, H., Chen, C. & Puett, D. *Endocrinology* **134,** 1768–1770 (1994).
31. Sairam, M. R. *FASEB J.* **3,** 1915–1926 (1989).
32. McFarland, K. C. *et al. Science* **245,** 494–499 (1989).
33. Chen, H-C., Shimohigashi, Y., Dufau, M. L. & Catt, K. J. *J. biol. Chem.* **257,** 14446–14452 (1982).
34. Sairam, M. R. & Bhargavi, G. N. *Science* **229,** 65–67 (1985).
35. Matzuk, M. M., Keene, J. L. & Boime, I. *J. biol. Chem.* **264,** 2409–2414 (1989).
36. Sairam, M. R. & Jiang, L. G. *Molec. cell. Endocr.* **85,** 227–235 (1992).
37. Rebois, R. V. & Liss, M. T. *J. biol. Chem.* **262,** 3891–3896 (1987).
38. Blithe, D. L. in *Glycoprotein Hormones* (eds Lustbader, J. W., Puett, D. & Ruddon, R. W.) 156–166 (Springer, New York, 1994).
39. Birken, S. *et al. Endocrinology* **123,** 572–583 (1988).
40. *CCP4* (SERC (UK) Collaborative Computing Project 4, Daresbury Lab. Warrington, UK, 1992).
41. Kabsch, W. *J. appl. Crystallogr.* **21,** 916–924 (1988).
42. Sheldrick, G. M. *Acta crystallogr.* **A46,** 467–473 (1990).
43. Wang, B. C. *Meth. Enzym.* **115,** 90–112 (1985).
44. Bricogne, G. & Gilmore, C. *Acta crystallogr.* **A46,** 284–297 (1990).
45. Read, R. J., *Acta crystallogr.* **A42,** 140–149 (1986).
46. Jones, T. A. *Meth. Enzym.* **115,** 157–171 (1985).
47. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta. crystallogr.* **A47,** 110–119 (1991).
48. Holmgren, A. & Morgan, F. *Eur. J. Biochem.* **70,** 377–383 (1976).
49. Nicholls, A., Sharp, K. A. & Honig, B. *Proteins* **11,** 281–296 (1991).
50. Jones, D. T., Taylor, W. R. & Thornton, J. M. *Compl. appl. Biosci.* **8,** 275–282 (1992).
51. Bernstein, F. C. *et al. J. molec. Biol.* **112,** 535–542 (1977).

ACKNOWLEDGEMENTS. We thank G. MacDermott and A. Freer for assistance with data collection; P. Emsley for help with computation; E. Dodson for advice on phase refinement; S. Birken for preparing some of the protein material; C. Carter and S. Xiang for advice on entropy calculations; and the staff at the SERC Daresbury Synchrotron Radiation Source for assistance. This work was supported by the MRC, in part by the WHO Task Force on Vaccines for Fertility Regulation and by a NATO Travel Award.

nature
INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
Volume 369 - No. 6480 - 9 June 1994 - $8.50



Tunable perovskites